

Rachel Geman
LIEFF CABRASER HEIMANN
& BERNSTEIN, LLP
250 Hudson Street, 8th Floor
New York, NY 10013-1413
Telephone: 212.355.9500
rgeman@lchb.com

Scott J. Sholder
CeCe M. Cole
COWAN DEBAETS ABRAHAM
& SHEPPARD LLP
41 Madison Avenue, 38th Floor
New York, New York 10010
Telephone: 212.974.7474
ssholder@cdas.com
ccole@cdas.com

*Attorneys for Plaintiffs and the Proposed Class
(Additional Counsel Listed on Signature Page)*

**UNITED STATES DISTRICT COURT
SOUTHERN DISTRICT OF NEW YORK**

AUTHORS GUILD, DAVID BALDACCI,
MARY BLY, MICHAEL CONNELLY, SYLVIA
DAY, JONATHAN FRANZEN, JOHN
GRISHAM, ELIN HILDERBRAND,
CHRISTINA BAKER KLINE, MAYA
SHANBHAG LANG, VICTOR LAVALLE,
GEORGE R.R. MARTIN, JODI PICOULT,
DOUGLAS PRESTON, ROXANA ROBINSON,
GEORGE SAUNDERS, SCOTT TUROW, and
RACHEL VAIL, individually and on behalf of
others similarly situated,

Plaintiffs,

v.

OPENAI INC., OPENAI OPCO LLC, OPENAI
GP LLC, OPENAI LLC, OPENAI GLOBAL
LLC, OAI CORPORATION LLC, OPENAI
HOLDINGS LLC, OPENAI STARTUP FUND I
LP, OPENAI STARTUP FUND GP I LLC,
OPENAI STARTUP FUND MANAGEMENT
LLC, and MICROSOFT CORPORATION,

Defendants.

No. 1:23-cv-8282-SHS

CLASS ACTION COMPLAINT

JURY TRIAL DEMANDED

INTRODUCTORY STATEMENT

1. Plaintiffs, authors of a broad array of works of fiction, bring this action under the Copyright Act seeking redress for Defendants' flagrant and harmful infringements of Plaintiffs'

registered copyrights in written works of fiction. Defendants copied Plaintiffs' works wholesale, without permission or consideration. Defendants then fed Plaintiffs' copyrighted works into their "large language models" or "LLMs," algorithms designed to output human-seeming text responses to users' prompts and queries. These algorithms are at the heart of Defendants' massive commercial enterprise. And at the heart of these algorithms is systematic theft on a mass scale.

2. Plaintiffs seek to represent a class of professional fiction writers whose works spring from their own minds and their creative literary expression. These authors' livelihoods derive from the works they create. But OpenAI's LLMs endanger fiction writers' ability to make a living, in that the LLMs allow anyone to generate—automatically and freely (or very cheaply)—texts that they would otherwise pay writers to create. Moreover, Open AI's LLMs can spit out derivative works: material that is based on, mimics, summarizes, or paraphrases Plaintiffs' works, and harms the market for them.

3. Unfairly, and perversely, without Plaintiffs' copyrighted works on which to "train" their LLMs, Defendants would have no commercial product with which to damage—if not usurp—the market for these professional authors' works. OpenAI's willful copying thus makes Plaintiffs' works into engines of their own destruction.

4. OpenAI could have "trained" its LLMs on works in the public domain. It could have paid a reasonable licensing fee to use copyrighted works. What Defendants could *not* do was evade the Copyright Act altogether to power their lucrative commercial endeavor, taking whatever datasets of relatively recent books they could get their hands on without authorization. There is nothing fair about this. OpenAI's unauthorized use of Plaintiffs' copyrighted works thus

presents a straightforward infringement case applying well-established law to well-recognized copyright harms.

5. OpenAI’s chief executive Sam Altman has told Congress that he shares Plaintiffs’ concerns. According to Altman, “Ensuring that the creator economy continues to be vibrant is an important priority for OpenAI. ... OpenAI does not want to replace creators. We want our systems to be used to empower creativity, and to support and augment the essential humanity of artists and creators.”¹ Altman testified that OpenAI “think[s] that creators deserve control over how their creations are used” and that “content creators, content owners, need to benefit from this technology.”² Altman also has represented that OpenAI has “licens[ed] content directly from content owners” for “training” purposes.³ Not so from Plaintiffs. As to them, Altman and Defendants have proved unwilling to turn these words into actions.

6. Plaintiffs thus seek damages for the lost opportunity to license their works, and for the market usurpation Defendants have enabled by making Plaintiffs unwilling accomplices in their own replacement; and a permanent injunction to prevent these harms from recurring.

7. Plaintiffs complain of Defendants, on personal knowledge as to matters relating to Plaintiffs themselves, and on information and belief based on their and their counsels’ reasonable investigation as to all other matters, as follows:

¹ Sam Altman, *Questions for the Record*, at 9–10 (June 22, 2023), available at https://www.judiciary.senate.gov/imo/media/doc/2023-05-16_-_qfr_responses_-_altman.pdf (last accessed Dec. 4, 2023).

² *Oversight of A.I.: Rules for Artificial Intelligence: Hearing Before the S. Judiciary Comm. Subcomm. on Privacy, Tech. and the Law*, 118th Cong. (2023) (testimony of OpenAI CEO Sam Altman), available at <https://techpolicy.press/transcript-senate-judiciary-subcommittee-hearing-on-oversight-of-ai> (last accessed Dec. 4, 2023).

³ Altman, *Questions for the Record*, *supra*, at 10.

JURISDICTION AND VENUE

8. The Court has jurisdiction over the subject matter of this action under 28 U.S.C. § 1338(a) because the action arises under the Copyright Act.

9. Venue is proper in this District under 28 U.S.C. § 1391(b)(2) because a substantial part of the events giving rise to Plaintiffs' claims occurred here.

10. Venue is also proper in this District under 28 U.S.C. § 1400(a) because Defendants or their agents reside or may be found here.

PARTIES

I. Plaintiffs

11. Plaintiff **The Authors Guild** is a nonprofit 501(c)(6) organization based in New York, New York.

12. Plaintiff **David Baldacci** is an author and a resident of Vienna, Virginia.

13. Plaintiff **Mary Bly** is an author and a resident of New York, New York.

14. Plaintiff **Michael Connelly** is an author and a resident of Tampa, Florida.

15. Plaintiff **Sylvia Day** is an author and a resident of Las Vegas, Nevada.

16. Plaintiff **Jonathan Franzen** is an author and a resident of Santa Cruz, California.

17. Plaintiff **John Grisham** is an author and a resident of Charlottesville, Virginia.

18. Plaintiff **Elin Hilderbrand** is an author and a resident of Nantucket Island, Massachusetts.

19. Plaintiff **Christina Baker Kline** is an author and a resident of New York, New York.

20. Plaintiff **Maya Shanbhag Lang** is an author and a resident of Sleepy Hollow, New York.

21. Plaintiff **Victor LaValle** is an author and a resident of New York, New York.

22. Plaintiff **George R.R. Martin** is an author and a resident of Santa Fe, New Mexico.

23. Plaintiff **Jodi Picoult** is an author and a resident of Hanover, New Hampshire.

24. Plaintiff **Douglas Preston** is an author and a resident of Santa Fe, New Mexico.

25. Plaintiff **Roxana Robinson** is an author and a resident of New York, New York.

26. Plaintiff **George Saunders** is an author and a resident of Santa Monica, California.

27. Plaintiff **Scott Turow** is an author and a resident of Naples, Florida.

28. Plaintiff **Rachel Vail** is an author and a resident of New York, New York.

II. **Defendants**

A. **OpenAI**

29. The Defendants described in paragraphs 31 through 48 below (collectively, “OpenAI” or the “OpenAI Defendants”) are a tangled thicket of interlocking entities that generally keep from the public what the precise relationships among them are and what function each entity serves within the larger corporate structure.

30. OpenAI is based in California and conducts its generative AI “training” from a facility in Iowa. OpenAI has caused injury in New York; has employees in New York; works with Microsoft in New York; and sells its products in New York.

31. Defendant **OpenAI Inc.** is a Delaware corporation with its principal place of business in San Francisco, California.

32. OpenAI Inc. was founded as a nonprofit research entity in 2015.

33. Defendant **OpenAI OpCo LLC** is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

34. Previously, OpenAI OpCo LLC was known as OpenAI LP, which was founded in 2019 to be the profit-making arm of OpenAI. OpenAI OpCo LLC now holds that role.

35. Defendant **OpenAI GP LLC** is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

36. OpenAI GP LLC is the vehicle through which OpenAI Inc. controls OpenAI OpCo LLC.

37. Defendant **OpenAI LLC** is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

38. OpenAI LLC owns some or all of the services and products provided by OpenAI.

39. The sole member of OpenAI LLC is Defendant OpenAI OpCo LLC.

40. Defendant **OpenAI Global LLC** is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

41. OpenAI Global's members are Defendant Microsoft Corporation and Defendant OAI Corporation LLC.

42. Defendant **OAI Corporation LLC** is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

43. OAI Corporation's only member is Defendant OpenAI Holdings LLC.

44. Defendant **OpenAI Holdings LLC** is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

45. The members of OpenAI Holdings LLC are Defendant OpenAI Inc. and Aestas LLC, an OpenAI-related limited liability company that is not a defendant here.

46. Defendant **OpenAI Startup Fund I LP** is a limited partnership formed under the laws of Delaware with its principal place of business in San Francisco, California.

47. Defendant **OpenAI Startup Fund GP I LLC** is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

48. Defendant **OpenAI Startup Fund Management LLC** is a limited liability company formed under the laws of Delaware with its principal place of business in San Francisco, California.

B. Microsoft

49. Defendant **Microsoft Corporation** is a Washington corporation with its principal place of business in Redmond, Washington.

50. It also has offices in New York, including an office and research lab in New York City. Microsoft's New York City-based operations "span[] a wide variety of topics within theoretical and applied machine learning, including learning from interactive data (e.g., contextual bandits and reinforcement learning), online learning, natural language processing, and topics related to interpretability and fairness of [machine learning] and [artificial intelligence]." ⁴

51. Specifically, Microsoft's Azure project, which collaborated heavily with OpenAI in its LLM "training," ⁵ does significant work in Microsoft's New York City office. ⁶

⁴ Microsoft, *Machine Learning & AI | NYC*, available at <https://www.microsoft.com/en-us/research/theme/machine-learning-ai-nyc/> (last accessed Dec. 4, 2023).

⁵ Jennifer Langston, *Microsoft announces new supercomputer, lays out vision for future AI work*, Microsoft (May 19, 2020), available at <https://news.microsoft.com/source/features/ai/openai-azure-supercomputer/> (last accessed Dec. 4, 2023).

⁶ John Roach, *Microsoft responsible machine learning capabilities build trust in AI systems, developers say*, Microsoft (May 19, 2020), available at <https://news.microsoft.com/source/features/ai/azure-responsible-machine-learning/> (last accessed Dec. 4, 2023).

52. Microsoft works in close partnership with OpenAI in connection with the LLMs at issue in this lawsuit. An OpenAI LLM underlies Microsoft's Bing Chat product, offered through its Bing search engine. Microsoft has also integrated OpenAI's LLMs into its sales and marketing software, coding tools, productivity software, and cloud storage services.⁷

53. Based on its \$13 billion investment in OpenAI and for other reasons, Microsoft has "significant rights" in OpenAI. As Microsoft, through its CEO, has explained, "We are below them [OpenAI], above them, around them."⁸

54. As CEO Satya Nadella further explained, "We have all the IP rights and all the capability. If OpenAI disappeared tomorrow, I don't want any customer of ours to be worried about it quite honestly, because we have all of the rights to continue the innovation. Not just to serve the product, but we can go and just do what we were doing in partnership ourselves. We have the people, we have the compute, we have the data, we have everything."⁹

55. Microsoft is estimated to hold a 49 percent stake in the "for-profit arm" of OpenAI.¹⁰

⁷ Jordan Novet, *Microsoft's \$13 billion bet on OpenAI carries huge potential along with plenty of uncertainty*, CNBC (Apr. 9, 2023), available at <https://www.cnbc.com/2023/04/08/microsofts-complex-bet-on-openai-brings-potential-and-uncertainty.html> (last accessed Dec. 3, 2023).

⁸ Intelligencer Staff, *Satya Nadella on Hiring the Most Powerful Man in AI When OpenAI threw Sam Altman overboard, Microsoft's CEO saw an opportunity*, Intelligencer (Nov. 21, 2023), available at <https://nymag.com/intelligencer/2023/11/on-with-kara-swisher-satya-nadella-on-hiring-sam-altman.html> (last accessed Dec. 1, 2023).

⁹ Intelligencer Staff, *Satya Nadella on Hiring the Most Powerful Man in AI When OpenAI threw Sam Altman overboard, Microsoft's CEO saw an opportunity*, Intelligencer (Nov. 21, 2023), available at <https://nymag.com/intelligencer/2023/11/on-with-kara-swisher-satya-nadella-on-hiring-sam-altman.html> (last accessed Dec. 1, 2023).

¹⁰ Julia Angwin, *The OpenAI Coup Is Great for Microsoft. What Does It Mean for Us?*, New York Times (Nov. 21, 2023), available at <https://www.nytimes.com/2023/11/21/opinion/the-sam-altman-openai-board-microsoft.html> (last accessed Dec. 4, 2023).

GENERAL FACTUAL ALLEGATIONS

I. Generative AI and Large Language Models

56. The terms “artificial intelligence” or “AI” refer generally to computer systems designed to imitate human cognitive functions.

57. The terms “generative artificial intelligence” or “generative AI” refer specifically to systems that are capable of generating “new” content in response to user inputs called “prompts.”

58. For example, the user of a generative AI system capable of generating images from text prompts might input the prompt, “A lawyer working at her desk.” The system would then attempt to construct the prompted image. Similarly, the user of a generative AI system capable of generating text from text prompts might input the prompt, “Tell me a story about a lawyer working at her desk.” The system would then attempt to generate the prompted text.

59. Recent generative AI systems designed to recognize input text and generate output text are built on “large language models” or “LLMs.”

60. LLMs use predictive algorithms that are designed to detect statistical patterns in the text datasets on which they are “trained” and, on the basis of these patterns, generate responses to user prompts. “Training” an LLM refers to the process by which the parameters that define an LLM’s behavior are adjusted through the LLM’s ingestion and analysis of large “training” datasets.

61. Once “trained,” the LLM analyzes the relationships among words in an input prompt and generates a response that is an approximation of similar relationships among words in the LLM’s “training” data. In this way, LLMs can be capable of generating sentences, paragraphs, and even complete texts, from cover letters to novels.

62. “Training” an LLM requires inputting large numbers of parameters in the model and then supplying the LLM with large amounts of text for the LLM to ingest—the more text, the better. That is, in part, the *large* in *large language model*.

63. As the U.S. Patent and Trademark Office has observed, LLM “training” “almost by definition involve[s] the reproduction of entire works or substantial portions thereof.”¹¹

64. “Training” in this context is therefore a technical-sounding euphemism for “copying and ingesting expression.”

65. Moreover, in some form and to some degree currently unknowable to the public, OpenAI’s LLMs have “memorized” or stored their “training” data (even if in a “translated” form), such that the data (at least in part) can be accessed, recalled, and reproduced by the LLM at will.¹²

66. The quality of the LLM (that is, its capacity to generate human-seeming responses to prompts) is dependent on the quality of the datasets used to “train” the LLM.

67. Professionally authored, edited, and published books—such as those authored by Plaintiffs here—are an especially important source of LLM “training” data.

68. As one group of AI researchers (not affiliated with Defendants) has observed, “[b]ooks are a rich source of both fine-grained information, how a character, an object or a scene

¹¹ U.S. Patent & Trademark Office, *Public Views on Artificial Intelligence and Intellectual Property Policy* 29 (2020), available at https://www.uspto.gov/sites/default/files/documents/USPTO_AI-Report_2020-10-07.pdf (last accessed Dec. 4, 2023).

¹² See Jason Koebler, *Google Researchers’ Attack Prompts ChatGPT to Reveal Its Training Data*, 404 Media (Nov. 29, 2023), available at <https://www.404media.co/google-researchers-attack-convinces-chatgpt-to-reveal-its-training-data/> (last accessed Dec. 4, 2023); Kent K. Chang *et al.*, *Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4* (2023), available at <https://arxiv.org/pdf/2305.00118v1.pdf> (last accessed Dec. 4, 2023).

looks like, as well as high-level semantics, what someone is thinking, feeling and how these states evolve through a story.”¹³

69. In other words, books are the high-quality materials Defendants want, need, and have therefore outright pilfered to develop generative AI products that produce high-quality results: text that appears to have been written by a human writer.

70. This use is highly commercial.

II. OpenAI’s Willful Infringement of Plaintiffs’ Copyrights

A. OpenAI

71. OpenAI (specifically, Defendant OpenAI Inc.) was founded in 2015 as a non-profit organization with the self-professed goal of researching and developing AI tools “unconstrained by a need to generate financial return.”¹⁴

72. Four years later, in 2019, OpenAI relaunched itself (specifically, through Defendant OpenAI GP LLC and Defendant OpenAI OpCo LLC)¹⁵ as a for-profit enterprise.

73. Investments began pouring in. Microsoft Corporation, one of the world’s largest technology companies, invested \$1 billion in 2019, an estimated \$2 billion in 2021, and a staggering \$10 billion in 2023, for a total investment of \$13 billion.

74. Industry observers currently value OpenAI at up to \$80 billion.¹⁶

¹³ Yukun Zhu *et al.*, *Aligning Books and Movies: Towards Story-like Visual Explanations by Watching Movies and Reading Books* 1 (2015), available at <https://arxiv.org/pdf/1506.06724.pdf> (last accessed Dec. 4, 2023).

¹⁴ OpenAI, *Introducing OpenAI* (Dec. 11, 2015), <https://openai.com/blog/introducing-openai> (last accessed Dec. 4, 2023).

¹⁵ Defendant OpenAI OpCo LLC was then known as OpenAI LP.

¹⁶ Kate Clark, *Thrive Capital to Lead Purchase of OpenAI Employee Shares at \$80 Billion-Plus Valuation*, *The Information* (Oct. 19, 2023), available at <https://www.theinformation.com/articles/thrive-capital-to-lead-purchase-of-openai-employee-shares-at-80-billion-plus-valuation> (last accessed Dec. 4, 2023).

B. GPT-N and ChatGPT

75. OpenAI's LLMs are collectively referred to as "GPT-N," which stands for "Generative Pre-trained Transformer" (a specific type of LLM architecture), followed by a version number.

76. GPT-3 was released in 2020 and exclusively licensed to Microsoft the same year.

77. OpenAI further refined GPT-3 into GPT-3.5, which was released in 2022.

78. In November 2022, OpenAI released ChatGPT, a consumer-facing chatbot application built on GPT-3.5.

79. ChatGPT's popularity exploded virtually overnight. By January 2023, less than three months after its release, the application had an estimated 100 million monthly active users, making it one of the fastest-growing consumer applications in history.

80. GPT-4, the successor to GPT-3.5, was released in March 2023.

81. GPT-4 underlies OpenAI's new subscription-based chatbot, called ChatGPT Plus, which is available to consumers for \$20 per month.

82. Defendants intend to earn billions of dollars from this technology.

83. When announcing the release of ChatGPT Enterprise, a subscription-based high-capability GPT-4 application targeted for corporate clients, in August 2023, OpenAI claimed that teams in "over 80% of Fortune 500 companies" were using its products.¹⁷

84. GPT-4 also underlies Microsoft's Bing Chat product, offered through its Bing Internet search engine, and is integrated into its sales and marketing software, coding tools, productivity software, and cloud storage services.

¹⁷ OpenAI, *Introducing ChatGPT Enterprise* (Aug. 28, 2023), <https://openai.com/blog/introducing-chatgpt-enterprise> (last accessed Dec. 4, 2023).

85. OpenAI CEO Sam Altman recently reported to OpenAI employees that OpenAI is on track to generate \$1.3 billion in revenue in 2023.¹⁸

86. Analysts estimate that Microsoft could earn more than \$10 billion in annual revenue by 2026 *only* from AI add-ons to its Microsoft 365 productivity software, “at the core” of which lies OpenAI technology.¹⁹

C. Knowingly “Training” GPT-N on Copyrighted Books

87. OpenAI does not disclose or publicize with specificity what datasets GPT-3, GPT-3.5, or GPT-4 were “trained” on. Despite its name, OpenAI treats that information as proprietary.

88. To “train” its LLMs—including GPT-3, GPT-3.5, and GPT-4—OpenAI has reproduced copyrighted books—including copyrighted books authored by Plaintiffs here—without their authors’ consent.

89. OpenAI has admitted as much.

90. OpenAI has admitted that it has “trained” its LLMs on “large, publicly available datasets that include copyrighted works.”²⁰

¹⁸ Amir Efrati, *OpenAI’s Revenue Crossed \$1.3 Billion Annualized Rate, CEO Tells Staff*, The Information (Oct. 12, 2023), available at <https://www.theinformation.com/articles/openai-revenue-crossed-1-3-billion-annualized-rate-ceo-tells-staff> (last accessed Dec. 4, 2023).

¹⁹ Jordan Novet, *Microsoft starts selling AI tool for Office, which could generate \$10 billion a year by 2026*, CNBC (Nov. 1, 2023), available at <https://www.cnbc.com/2023/11/01/microsoft-365-copilot-becomes-generally-available.html> (last accessed Dec. 4, 2023).

²⁰ OpenAI, *Comment Regarding Request for Comments on Intellectual Property Protection for Artificial Intelligence Innovation*, U.S. Patent and Trademark Office Dkt. No. PTO-C-2019-0038, at 1 (2019), available at https://www.uspto.gov/sites/default/files/documents/OpenAI_RFC-84-FR-58141.pdf (last accessed Dec. 4, 2023).

91. Again: OpenAI’s “training” data is “derived from existing publicly accessible ‘corpora’ ... of data that include copyrighted works.”²¹

92. OpenAI has admitted that “training” LLMs “require[s] large amounts of data,” and that “analyzing large corpora” of data “necessarily involves first making copies of the data to be analyzed.”²²

93. OpenAI has admitted that, if it refrained from using copyrighted works in its LLMs’ “training,” it would “lead to significant reductions in model quality.”²³

94. Accordingly, OpenAI has openly admitted to reproducing copyrighted works in the course of “training” its LLMs because such reproduction is central to the quality of its products.

95. ChatGPT itself admits as much. In response to a query submitted to it in January 2023, the chatbot responded,

It is possible that some of the books used to train me were under copyright. However, my training data was sourced from various publicly available sources on the internet, and it is likely that some of the books included in my training dataset were not authorized to be used. ... If any copyrighted material was included in my training data, it would have been used without the knowledge or consent of the copyright holder.

96. Until very recently, ChatGPT could be prompted to return quotations of text from copyrighted books with a good degree of accuracy, suggesting that the underlying LLM must have ingested these books in their entirety during its “training.”

²¹ *Id.* at 2.

²² *Id.*

²³ *Id.* at 7 n.33.

97. Now, however, ChatGPT generally responds to such prompts with the statement, “I can’t provide verbatim excerpts from copyrighted texts.” Thus, while ChatGPT previously provided such excerpts and in principle retains the capacity to do so, it has been restrained from doing so, if only temporarily, by its programmers.

98. In light of its timing, this apparent revision of ChatGPT’s output rules is likely a response to the type of activism on behalf of authors exemplified by the Open Letter addressed to OpenAI and other companies by Plaintiff The Authors Guild, which is discussed further below.

99. Instead of “verbatim excerpts,” ChatGPT now offers to produce a summary of the copyrighted book, which usually contains details not available in reviews and other publicly available material—again suggesting that the underlying LLM must have ingested the entire book during its “training.”

100. OpenAI is characteristically opaque about where and how it procured the entirety of these books, including Plaintiffs’ copyrighted works.

101. OpenAI has discussed limited details about the datasets used to “train” GPT-3.

102. OpenAI admits that among the “training” datasets it used to “train” the model were “Common Crawl,” and two “high-quality,” “internet-based books corpora” which it calls “Books1” and “Books2.”²⁴

103. Common Crawl is a vast and growing corpus of “raw web page data, metadata extracts, and text extracts” scraped from billions of web pages. It is widely used in “training”

²⁴ Tom B. Brown *et al.*, *Language Models Are Few-Shot Learners* 8 (2020), available at <https://arxiv.org/pdf/2005.14165.pdf> (last accessed Dec. 4, 2023).

LLMs, and has been used to “train,” in addition to GPT-N, Meta’s LLaMa, and Google’s BERT.

It is known to contain text from books copied from pirate sites.²⁵

104. OpenAI refuses to discuss the source or sources of the Books2 dataset.

105. Some independent AI researchers suspect that Books2 contains or consists of ebook files downloaded from large pirate book repositories such as Library Genesis or “LibGen,” “which offers a vast repository of pirated text.”²⁶

106. LibGen is already known to this Court as a notorious copyright infringer.²⁷

107. Other possible candidates for Books2’s sources include Z-Library, another large pirate book repository that hosts more than 11 million books, and pirate torrent trackers like Bibliotik, which allow users to download ebooks in bulk.

108. Websites linked to Z-Library appear in the Common Crawl corpus and have been included in the “training” dataset of other LLMs.²⁸

109. Z-Library’s Internet domains were seized by the FBI in February 2022, only months after OpenAI stopped “training” GPT-3.5 in September 2021.

²⁵ Alex Hern, *Fresh Concerns Raised Over Sources of Training Material for AI Systems*, The Guardian (Apr. 20, 2023), available at <https://www.theguardian.com/technology/2023/apr/20/fresh-concerns-training-material-ai-systems-facist-pirated-malicious> (last accessed Dec. 4, 2023).

²⁶ Kate Knibbs, *The Battle Over Books3 Could Change AI Forever*, Wired (Sept. 4, 2023), available at <https://www.wired.com/story/battle-over-books3> (last accessed Dec. 4, 2023).

²⁷ See *Elsevier Inc. v. Sci-Hub*, No. 1:15-cv-4282-RWS (S.D.N.Y.).

²⁸ Kevin Schaul *et al.*, *Inside the Secret List of Websites that Make AI Like ChatGPT Sounds Smart*, The Washington Post (Apr. 19, 2023), available at <https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning> (last accessed Dec. 4, 2023).

110. The disclosed size of the Books2 dataset (55 billion “tokens,” the basic units of textual meaning such as words, syllables, numbers, and punctuation marks) suggests it comprises over 100,000 books.

111. “Books3,” a dataset compiled by an independent AI researcher, is comprised of nearly 200,000 books downloaded from Bibliotik, and has been used by other AI developers to “train” LLMs.

112. The similarities in the sizes of Books2 and Books3, and the fact that there are only a few pirate repositories on the Internet that allow bulk ebook downloads, strongly indicates that the books contained in Books2 were also obtained from one of the notorious repositories discussed above.

113. OpenAI has not discussed the datasets used to “train” GPT-3.5, GPT-4, or their source or sources.

114. GPT-3.5 and GPT-4 are significantly more powerful than their predecessors. GPT 3.5 contains roughly 200 billion parameters, and GPT 4 contains roughly 1.75 trillion parameters, compared to GPT-3’s roughly 175 billion parameters.

115. The growth in power and sophistication from GPT-3 to GPT-4 suggests a correlative growth in the size of the “training” datasets, raising the inference that one or more very large sources of pirated ebooks discussed above must have been used to “train” GPT-4.

116. There is no other way OpenAI could have obtained the volume of books required to “train” a powerful LLM like GPT-4.

117. In short, OpenAI admits it needs²⁹ and uses³⁰ “large, publicly available datasets that include copyrighted works”³¹—and specifically, “high-quality”³² copyrighted books—to “train” its LLMs; pirated sources of such “training” data are readily available; and one or more of these sources contain Plaintiffs’ works.

118. Defendants knew that their “training” data included texts protected by copyright but willfully proceeded without obtaining authorization.

119. OpenAI’s “training” its LLMs could not have happened without Microsoft’s financial and technical support. In 2020, Microsoft announced that it had developed “one of the top five publicly disclosed supercomputers in the world” that had been “[b]uilt in collaboration with and exclusively for OpenAI,” and “designed specifically to train that company’s AI models.”³³ And in 2023, Microsoft CEO Satya Nadella reminded the world that the “heavy lifting” for OpenAI’s LLM “training” was done by Microsoft “compute infrastructure.”³⁴

²⁹ OpenAI, *Comment Regarding Request for Comments*, *supra*, at 7 n.33.

³⁰ *Id.* at 2.

³¹ *Id.* at 1.

³² Brown *et al.*, *Few-Shot Learners*, *supra*, at 8.

³³ Jennifer Langston, *Microsoft announces new supercomputer, lays out vision for future AI work*, Microsoft (May 19, 2020), available at <https://news.microsoft.com/source/features/ai/openai-azure-supercomputer/> (last accessed Dec. 4, 2023).

³⁴ CNBC, *First on CNBC: CNBC Transcript: Microsoft CEO Satya Nadella Speaks with CNBC’s Jon Fortt on “Power Lunch” Today* (Feb. 7, 2023), available at <https://www.cnbc.com/2023/02/07/first-on-cnbc-cnbc-transcript-microsoft-ceo-satya-nadella-speaks-with-cnbc-jon-fortt-on-power-lunch-today.html> (last accessed Dec. 4, 2023).

D. GPT-N's and ChatGPT's Harm to Authors

120. ChatGPT and the LLMs underlying it seriously threaten the livelihood of the very authors—including Plaintiffs here, as discussed specifically below—on whose works they were “trained” without the authors’ consent.

121. Goldman Sachs estimates that generative AI could replace 300 million full-time jobs in the near future, or one-fourth of the labor currently performed in the United States and Europe.

122. Already, writers report losing income from copywriting, journalism, and online content writing—important sources of income for many book authors. The Authors Guild’s most recent authors earnings study³⁵ shows a median writing-related income for full-time authors of just over \$20,000, and that full-time traditional authors earn only half of that from their books. The rest comes from activities like content writing—work that is starting to dry up as a result of generative AI systems like ChatGPT.

123. An Authors Guild member who writes marketing and web content reported losing 75 percent of their work as a result of clients switching to AI.

124. Another content writer (unrelated to the Plaintiffs here) told the *Washington Post* that half of his annual income (generated by ten client contracts) was erased when the clients elected to use ChatGPT instead.³⁶

³⁵ Authors Guild, “Top Takeaways from the 2023 Author Income Survey (2023),” <https://authorsguild.org/news/top-takeaways-from-2023-author-income-survey> (last accessed Dec. 4, 2023).

³⁶ Pranshu Verma & Gerrit De Vynck, *ChatGPT Took Their Jobs. Now They Walk Dogs and Fix Air Conditioners*, The Washington Post (June 2, 2023), available at <https://www.washingtonpost.com/technology/2023/06/02/ai-taking-jobs> (last accessed Dec. 4, 2023).

125. Recently, the owner of popular online publications such as *Gizmodo*, *Deadspin*, *The Root*, *Jezebel* and *The Onion* came under fire for publishing an error-riddled, AI-generated piece, leading the Writers Guild of America to demand “an immediate end of AI-generated articles” on the company’s properties.³⁷

126. In a survey of authors conducted by The Authors Guild in March 2023 (early in ChatGPT’s lifecycle), 69 percent of respondents said they consider generative AI a threat to their profession, and 90 percent said they believe that writers should be compensated for the use of their work in “training” AI.

127. As explained above, until recently, ChatGPT provided verbatim quotes of copyrighted text. Currently, it instead readily offers to produce summaries of such text. These summaries are themselves derivative works, the creation of which is inherently based on the original unlawfully copied work and could be—but for ChatGPT—licensed by the authors of the underlying works to willing, *paying* licensees.

128. ChatGPT creates other outputs that are derivative of authors’ copyrighted works. Businesses are sprouting up to sell prompts that allow users to enter the world of an author’s books and create derivative stories within that world. For example, a business called Socialdraft offers long prompts that lead ChatGPT to engage in “conversations” with popular fiction authors like Plaintiff Grisham, Plaintiff Martin, Margaret Atwood, Dan Brown, and others about their works, as well as prompts that promise to help customers “Craft Bestselling Books with AI.”

³⁷ Todd Spangler, *WGA Slams G/O Media’s AI-Generated Articles as ‘Existential Threat to Journalism,’ Demands Company End Practice*, Variety (July 12, 2023), <https://variety.com/2023/digital/news/wga-slams-go-media-ai-generated-articles-existential-threat-1235668496> (last accessed Dec. 4, 2023).

129. OpenAI allows third parties to build their own applications on top of ChatGPT by making it available through an “application programming interface” or “API.” Applications integrated with the API allow users to generate works of fiction, including books and stories similar to those of Plaintiffs and other authors.³⁸

130. ChatGPT is being used to generate low-quality ebooks, impersonating authors, and displacing human-authored books.³⁹ For example, author Jane Friedman discovered “a cache of garbage books” written under her name for sale on Amazon.⁴⁰

131. Plaintiffs and other professional writers are thus reasonably concerned about the risks OpenAI’s conduct poses to their livelihoods specifically and the literary arts generally.

132. Plaintiff The Authors Guild, among others, has given voice to these concerns on behalf of working American authors.

133. The Authors Guild is the nation’s oldest and largest professional writers’ organization. It “exists to support working writers and their ability to earn a living from authorship.”⁴¹

³⁸ Adi Robertson, *I Tried the AI Novel-Writing Tool Everyone Hates, and It’s Better than I Expected*, The Verge (May 24, 2023), <https://www.theverge.com/2023/5/24/23732252/sudowrite-story-engine-ai-generated-cyberpunk-novella> (last accessed Dec. 4, 2023).

³⁹ Jules Roscoe, *AI-Generated Books of Nonsense Are All Over Amazon’s Bestseller Lists*, Vice (June 28, 2023), <https://www.vice.com/en/article/v7b774/ai-generated-books-of-nonsense-are-all-over-amazons-bestseller-lists> (last accessed Dec. 4, 2023).

⁴⁰ Pilar Melendez, *Famous Author Jane Friedman Finds AI Fakes Being Sold Under Her Name on Amazon*, The Daily Beast (Aug. 8, 2023), <https://www.thedailybeast.com/author-jane-friedman-finds-ai-fakes-being-sold-under-her-name-on-amazon> (last accessed Dec. 4, 2023).

⁴¹ Authors Guild, <https://authorsguild.org> (last accessed Dec. 4, 2023).

134. Among other principles, The Authors Guild holds that “authors should not be required to write or speak without compensation. Writers, like all professionals, should receive fair payment for their work.”⁴²

135. In June 2023, The Authors Guild wrote an open letter (the “Open Letter”) calling on OpenAI and other major technology companies to fairly license authors’ works for use in LLM “training.”

136. The Open Letter emphasizes that “[g]enerative AI technologies built on large language models owe their existence to our writings,” and protests “the inherent injustice in exploiting our works as part of your AI systems without our consent, credit, or compensation.”⁴³

137. The Open Letter also points to the risks to authors’ livelihoods posed by generative AI like GPT-N and ChatGPT: “As a result of embedding our writings in your systems, generative AI threatens to damage our profession by flooding the market with mediocre, machine-written books, stories, and journalism based on our work. ... The introduction of generative AI threatens ... to make it even more difficult, if not impossible, for writers—especially young writers and voices from under-represented communities—to earn a living from their profession.”⁴⁴

⁴² Authors Guild, *Principles*, <https://authorsguild.org/about/principles> (last accessed Dec. 4, 2023).

⁴³ Open Letter from The Authors Guild to Sam Altman *et al.*, at 1, *available at* <https://authorsguild.org/app/uploads/2023/07/Authors-Guild-Open-Letter-to-Generative-AI-Leaders.pdf> (last accessed Dec. 4, 2023).

⁴⁴ *Id.*

138. To date, the Open Letter has been signed by almost 12,000 authors,⁴⁵ including many Plaintiffs here.⁴⁶

139. In short, the success and profitability of OpenAI are predicated on mass copyright infringement without a word of permission from or a nickel of compensation to copyright owners, including Plaintiffs here. OpenAI knows it; its investors know it; and Plaintiffs know it.

PLAINTIFF-SPECIFIC ALLEGATIONS

140. Plaintiffs' works collectively span a wide range of commercial fiction whose continuing commercial viability is endangered by Defendants. Each author represented here has a distinct voice, a distinct style, and distinct creative expression. But all Plaintiffs have suffered identical harms from OpenAI's infringing reproductions of their works.

141. The contents of the datasets OpenAI has used to "train" its LLMs are peculiarly within its knowledge and not publicly disclosed, such that Plaintiffs are unable discern those contents with perfect accuracy. Plaintiffs make the specific allegations of infringement below based on what is known about OpenAI's training practices; what is known about the contents, uses, and availability of the pirate book repositories such as LibGen, Bibliotik, and Z-Library; and the results of Plaintiffs' testing of ChatGPT.

142. Most Plaintiffs have written more books than are included in this Complaint.

I. Plaintiff The Authors Guild

143. The Authors Guild is the owner of the registered copyrights in Mignon Eberhart's works, including *While the Patient Slept* and *The Patient in Room 18*.

⁴⁵ Authors Guild, *Open Letter to Generative AI Leaders*, <https://actionnetwork.org/petitions/authors-guild-open-letter-to-generative-ai-leaders> (last accessed Dec. 4, 2023).

⁴⁶ See Open Letter, *supra*, at 2–124.

144. Mignon G. Eberhart (1899–1996), dubbed “America’s Agatha Christie,” was the author of dozens of mystery novels over nearly sixty years. Several of Eberhart’s novels have been adapted for film, including *Hasty Wedding*, *Mystery House*, *While the Patient Slept*, *The Patient in Room 18*, and *The White Cockatoo*.

145. The Authors Guild is the owner or beneficial owner of the registered copyrights in eleven (11) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Authors Guild Infringed Works”).

146. The registration information for the Authors Guild Infringed Works is contained in Exhibit A to this Complaint, at 1.

147. OpenAI unlawfully and willfully copied the Authors Guild Infringed Works and used them to “train” OpenAI’s LLMs without The Authors Guild’s permission.

148. For example, when prompted, ChatGPT accurately generated summaries of several of the Authors Guild Infringed Works, including summaries for *While the Patient Slept* and *The Patient in Room 18*.

149. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *While the Patient Slept*, one of the Authors Guild Infringed Works, and titled the infringing and unauthorized derivative “Shadows Over Federie House,” using the same characters from Eberhart’s existing book.

150. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Patient in Room 18*, one of the Authors Guild Infringed Works, and titled the infringing and unauthorized derivative “Echoes from Room 18,” using the same characters from Eberhart’s existing book.

151. When prompted, ChatGPT generated an accurate summary of the final chapter of *While the Patient Slept*, one of the Authors Guild Infringed Works.

152. ChatGPT could not have generated the material described above if OpenAI's LLMs had not ingested and been "trained" on the Authors Guild Infringed Works.

II. Plaintiff Baldacci

153. Baldacci is a best-selling author, philanthropist, and lawyer whose novels have been adapted for film and television, published in over 45 languages and in more than 80 countries, with 150 million copies sold worldwide. Some of Baldacci's most popular works include books in the *Camel Club* series, *Vega Jane* series, and *Archer* series.

154. Baldacci is a member of The Authors Guild.

155. Baldacci is the sole author of and owner or beneficial owner of the registered copyrights in forty-one (41) written works of fiction, all or many of which OpenAI ingested and copied without permission (the "Baldacci Infringed Works").

156. The registration information for the Baldacci Infringed Works is contained in Exhibit A to this Complaint, at 1–2.

157. OpenAI unlawfully and willfully copied the Baldacci Infringed Works and used them to "train" OpenAI's LLMs without Baldacci's permission.

158. For example, when prompted, ChatGPT accurately generated summaries of several of the Baldacci Infringed Works, including summaries of *The Collectors*, *The Finisher*, and *One Good Deed*.

159. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Simple Truth*, one of the Baldacci Infringed Works, and titled the infringing and unauthorized derivative "The Complex Justice," using the same characters from Baldacci's existing book.

160. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Total Control*, one of the Baldacci Infringed Works, and titled the infringing and unauthorized derivative “Total Control: Unfinished Business,” using the same characters from Baldacci’s existing book.

161. When prompted, ChatGPT generated an accurate summary of the final chapter of *Long Road to Mercy*, one of the Baldacci Infringed Works.

162. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Baldacci Infringed Works.

III. Plaintiff Bly

163. Bly is a tenured professor and chair of the English department at Fordham University who also writes best-selling Regency and Georgian romance novels under the pen name Eloisa James. Some of Bly’s most popular works include books in the *Desperate Duchesses* series, the *Fairy Tales* series, the *Wildes of Lindow Castle* series, and the *Essex* series.

164. Bly is a Vice President of The Authors Guild Council and a member of The Authors Guild.

165. Bly is the sole author of and owner or beneficial owner of the registered copyrights in thirty-three (33) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Bly Infringed Works”).

166. The registration information for the Bly Infringed Works is contained in Exhibit A to this Complaint, at 2–3.

167. OpenAI unlawfully and willfully copied the Bly Infringed Works used them to “train” OpenAI’s LLMs without Bly’s permission.

168. For example, when prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *This Duchess of Mine*, one of the Bly

Infringed Works, and titled the infringing and unauthorized derivative “The Duchess’ New Dawn,” using the same characters from Bly’s existing book.

169. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *An Affair Before Christmas*, one of the Bly Infringed Works, and titled the infringing and unauthorized derivative “Whispers of Winter,” using the same characters from Bly’s existing book.

170. When prompted, ChatGPT generated an accurate summary of the final chapter of *A Duke of Her Own*, one of the Bly Infringed Works.

171. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Bly Infringed Works.

IV. Plaintiff Connelly

172. Connelly is a best-selling author with over 85 million copies of his books sold worldwide and translated into 45 foreign languages. Some of Connelly’s most popular novels include *The Lincoln Lawyer*, *City of Bones*, and *The Law of Innocence*.

173. Connelly is a member of The Authors Guild.

174. Connelly is the sole author of and owner or beneficial owner of the registered copyrights in forty-six (46) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Connelly Infringed Works”).

175. The registration information for the Connelly Infringed Works is contained in Exhibit A to this Complaint, at 3–4.

176. OpenAI unlawfully and willfully copied the Connelly Infringed Works and used them to “train” OpenAI’s LLMs without Connelly’s permission.

177. For example, when prompted, ChatGPT accurately generated summaries of several of the Connelly Infringed Works, including summaries for *The Black Echo*, *The Poet*, and *The Crossing*.

178. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Lincoln Lawyer*, one of the Connelly Infringed Works, and titled the infringing and unauthorized derivative “The City’s Shadows,” using the same characters from Connelly’s existing book.

179. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Brass Verdict*, one of the Connelly Infringed Works, and titled the infringing and unauthorized derivative “Double-Edged Justice,” using the same characters from Connelly’s existing book.

180. When prompted, ChatGPT generated an accurate summary of the final chapter of *The Late Show*, one of the Connelly Infringed Works.

181. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Connelly Infringed Works.

V. Plaintiff Day

182. Day is a best-selling author of over twenty award-winning novels, including ten *New York Times* best sellers and thirteen *USA Today* best sellers. Her work has been translated into forty-one languages. Some of Day’s most popular novels include books in *The Crossfire*® *Saga* series, the *Georgian* series, and the *Marked* series.

183. Day is a member of The Authors Guild Council and a member of The Authors Guild.

184. Day is the sole author of and owner or beneficial owner of the registered copyrights in thirty-one (31) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Day Infringed Works”).

185. The registration information for the Day Infringed Works is contained in Exhibit A to this Complaint, at 4.

186. OpenAI unlawfully and willfully copied the Day Infringed Works and used them to “train” OpenAI’s LLMs without Day’s permission.

187. For example, when prompted, ChatGPT accurately generated summaries of several of the Day Infringed Works, including summaries for *Bared to You*, *One With You*, and *Ask For It*.

188. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *A Touch of Crimson*, one of the Day Infringed Works, and titled the infringing and unauthorized derivative “Crimson Temptations: A Love Rekindled,” using the same characters from Day’s existing book.

189. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Butterfly in Frost*, one of the Day Infringed Works, and titled the infringing and unauthorized derivative “Butterfly in Frost: Embers of Desire,” using the same characters from Day’s existing book.

190. When prompted, ChatGPT generated an accurate summary of the final chapter of *The Stranger I Married*, one of the Day Infringed Works.

191. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Day Infringed Works.

VI. Plaintiff Franzen

192. Franzen is a novelist whose honors include the National Book Award, the James Tait Black Memorial Award, the Heartland Prize, Die Welt Literature Prize, the Budapest Grand Prize, and the first Carlos Fuentes Medal awarded at the Guadalajara International Book Fair.

Franzen is a member of the American Academy of Arts and Letters, the American Academy of Arts and Sciences, the German Akademie der Künste, and the French Ordre des Arts et des Lettres. Some of Franzen's most popular novels include *The Corrections*, *Purity*, and *Freedom*.

193. Franzen is a member of The Authors Guild.

194. Franzen is the sole author of and owner or beneficial owner of the registered copyrights in five (5) written works of fiction, all or many of which OpenAI ingested and copied without permission (the "Franzen Infringed Works").

195. The registration information for the Franzen Infringed Works is contained in Exhibit A to this Complaint, at 4–5.

196. OpenAI unlawfully and willfully copied the Franzen Infringed Works and used them to "train" OpenAI's LLMs without Franzen's permission.

197. For example, when prompted, ChatGPT accurately generated summaries of several of the Franzen Infringed Works, including summaries for *The Corrections*, *Purity*, and *Freedom*.

198. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Corrections*, one of the Franzen Infringed Works, and titled the infringing and unauthorized derivative "Revisions," using the same characters from Franzen's existing book.

199. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Twenty-Seventh City*, one of the Franzen

Infringed Works, and titled the infringing and unauthorized derivative “The Rising Metropolis,” using the same characters from Franzen’s existing book.

200. When prompted, ChatGPT generated an accurate summary of the final chapter of *Freedom*, one of the Franzen Infringed Works.

201. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Franzen Infringed Works.

VII. Plaintiff Grisham

202. Grisham is a civically engaged and best-selling author. His award-winning work has been translated into approximately 50 languages and adapted for both television and film. Some of Grisham’s most popular novels include *The Pelican Brief*, *The Runaway Jury*, and *The Rainmaker*.

203. Grisham is a member of The Authors Guild.

204. Grisham is the sole author of and owner or beneficial owner of the registered copyrights in twenty-six (26) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Grisham Infringed Works”).

205. The registration information for the Grisham Infringed Works is contained in Exhibit A to this Complaint, at 5.

206. OpenAI unlawfully and willfully copied the Grisham Infringed Works and used them to “train” OpenAI’s LLMs without Grisham’s permission.

207. For example, when prompted, ChatGPT accurately generated summaries of several of the Grisham Infringed Works, including summaries for *The Chamber*, *The Client*, and *The Firm*.

208. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The King of Torts*, one of the Grisham Infringed

Works, and titled the infringing and unauthorized derivative “The Kingdom of Consequences,” using the same characters from Grisham’s existing book.

209. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Last Juror*, one of the Grisham Infringed Works, and titled the infringing and unauthorized derivative “The Juror’s Dilemma,” using the same characters from Grisham’s existing book.

210. When prompted, ChatGPT generated an accurate summary of the final chapter of *The Litigators*, one of the Grisham Infringed Works.

211. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Grisham Infringed Works.

VIII. Plaintiff Hilderbrand

212. Hilderbrand is a best-selling author, whose works include novels in the romance genre adapted for television. Hilderbrand has previously taught writing at the University of Iowa. Some of Hilderbrand’s most popular novels include *The Summer of ‘69*, *The Identicals*, and *The Perfect Couple*.

213. Hilderbrand is the sole author of and owner or beneficial owner of the registered copyrights in twenty-nine (29) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Hilderbrand Infringed Works”).

214. The registration information for the Hilderbrand Infringed Works is contained in Exhibit A to this Complaint, at 5–6.

215. OpenAI unlawfully and willfully copied the Hilderbrand Infringed Works and used them “train” OpenAI’s LLMs without Hilderbrand’s permission.

216. For example, when prompted, ChatGPT accurately generated summaries of several of the Hilderbrand Infringed Works, including summaries for *The Summer of '69*, *The Identicals*, and *The Perfect Couple*.

217. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Identicals*, one of the Hilderbrand Infringed Works, and titled the infringing and unauthorized derivative “The Reckoning of Twins,” using the same characters from Hilderbrand’s existing book.

218. When prompted, ChatGPT generated an accurate summary of the final chapter of *The Perfect Couple*, one of the Hilderbrand Infringed Works.

219. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Hilderbrand Infringed Works.

IX. Plaintiff Kline

220. Kline is a globally published author who writes best-selling novels and has taught different disciplines of writing at Yale University, New York University, and the University of Virginia. Some of Kline’s most popular novels include *Orphan Train*, *A Piece of the World*, and *Bird in Hand*.

221. Kline is a member of The Authors Guild Council and a member of The Authors Guild.

222. Kline is the sole author of and owner or beneficial owner of the registered copyrights in five (5) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Kline Infringed Works”).

223. The registration information for the Kline Infringed Works is contained in Exhibit A to this Complaint, at 6.

224. OpenAI unlawfully and willfully copied the Kline Infringed Works and used them to “train” OpenAI’s LLMs without Kline’s permission.

225. For example, when prompted, ChatGPT accurately generated summaries of several of the Kline Infringed Works, including summaries for *Orphan Train*, *A Piece of the World*, and *Bird in Hand*.

226. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Orphan Train*, one of the Kline Infringed Works, and titled the infringing and unauthorized derivative “Legacy Rails,” using the same characters from Kline’s existing book.

227. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Bird in Hand*, one of the Kline Infringed Works, and titled the infringing and unauthorized derivative “Ties That Bind,” using the same characters from Kline’s existing book.

228. When prompted, ChatGPT generated an accurate summary of the final chapter of *A Piece of the World*, one of the Kline Infringed Works.

229. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Kline Infringed Works.

X. Plaintiff Lang

230. Lang is an author and teacher who holds a doctorate in Comparative Literature. Lang is the author of the novel *The Sixteenth of June*.

231. Lang is the President and a member of The Authors Guild.

232. Lang is the sole author of and owner or beneficial owner of the registered copyrights in one (1) written work of fiction that OpenAI ingested and copied without permission (the “Lang Infringed Work”).

233. The registration information for the Lang Infringed Work is contained in Exhibit A to this Complaint, at 6.

234. OpenAI unlawfully and willfully copied the Lang Infringed Work and used it to “train” OpenAI’s LLMs without Lang’s permission.

235. When prompted, ChatGPT accurately generated a summary of the Lang Infringed Work, *The Sixteenth of June*.

236. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Sixteenth of June*, the Lang Infringed Work, and titled the infringing and unauthorized derivative “The Seventeenth of June,” using the same characters from Lang’s existing book.

237. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Lang Infringed Work.

XI. Plaintiff LaValle

238. LaValle is an associate professor of Creative Writing at Columbia University and the author of five novels, a short story collection, two novellas, and two comic books. Some of LaValle’s most popular novels include *Big Machine*, *The Devil in Silver*, and *The Changeling*.

239. LaValle is the sole author of and owner or beneficial owner of the registered copyrights in six (6) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “LaValle Infringed Works”).

240. The registration information for the LaValle Infringed Works is contained in Exhibit A to this Complaint, at 6.

241. OpenAI unlawfully and willfully copied the LaValle Infringed Works and used them to “train” OpenAI’s LLMs without LaValle’s permission.

242. For example, when prompted, ChatGPT accurately generated summaries of several of the LaValle Infringed Works, including summaries for *Big Machine*, *The Devil in Silver*, and *The Changeling*.

243. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Changeling*, one of the LaValle Infringed Works, and titled the infringing and unauthorized derivative “The Fae’s Return,” using the same characters from LaValle’s existing book.

244. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Devil in Silver*, one of the LaValle Infringed Works, and titled the infringing and unauthorized derivative “The New Hyde Legacy,” using the same characters from LaValle’s existing book.

245. When prompted, ChatGPT generated an accurate summary of the final chapter of *Big Machine*, one of the LaValle Infringed Works.

246. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the LaValle Infringed Works.

XII. Plaintiff Martin

247. Martin is an award-winning author, television producer, and writer who is widely known for his fantasy, science fiction, and horror writing. Some of Martin’s most popular novels include *A Game of Thrones*, *A Clash of Kings*, and *A Storm of Swords*.

248. Martin is the sole author of and owner or beneficial owner of the registered copyrights in fifteen (15) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Martin Infringed Works”).

249. The registration information for the Martin Infringed Works is contained in Exhibit A to this Complaint, at 6–7.

250. OpenAI unlawfully and willfully copied the Martin Infringed Works and used them to “train” OpenAI’s LLMs without Martin’s permission.

251. In July 2023, Liam Swayne used ChatGPT to generate versions of *The Winds of Winter* and *A Dream of Spring*, intended to be the final two books in the series *A Song of Ice and Fire*, which Martin is currently writing.

252. An experiment conducted by researchers at the University of California, Berkeley, into the “memorization” of works by ChatGPT found that Martin’s novel *A Game of Thrones* ranked 12th with respect to the degree of “memorization.”⁴⁷

253. When prompted, ChatGPT accurately generated summaries of several of the Martin Infringed Works, including summaries for Martin’s novels *A Game of Thrones*, *A Clash of Kings*, and *A Storm of Swords*, the first three books in the series *A Song of Ice and Fire*.

254. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for an alternate sequel to *A Clash of Kings*, one of the Martin Infringed Works, and titled the infringing and unauthorized derivative “A Dance With Shadows,” using the same characters from Martin’s existing books in the series *A Song of Ice and Fire*.

255. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for a prequel book to *A Game of Thrones*, one of the Martin Infringed Works, and titled the infringing and unauthorized derivative “A Dawn of Direwolves,” using the same characters from Martin’s existing books in the series *A Song of Ice and Fire*.

256. When prompted, ChatGPT generated an accurate summary of the final chapter of *The Armageddon Rag*, one of the Martin Infringed Works.

⁴⁷ See Kent K. Chang *et al.*, *Speak, Memory: An Archaeology of Books Known to ChatGPT/GPT-4* (2023), available at <https://arxiv.org/pdf/2305.00118v1.pdf> (last accessed Dec. 4, 2023).

257. ChatGPT could not have generated the results described above if OpenAI’s LLMs had not ingested and been “trained” on the Martin Infringed Works.

XIII. Plaintiff Picoult

258. A *New York Times* best-selling author, Picoult writes popular fiction. Picoult is also the recipient of many awards, including the New England Bookseller Award for Fiction, the Alex Awards from the YALSA, a lifetime achievement award for mainstream fiction from the Romance Writers of America, the NH Literary Award for Outstanding Literary Merit and the Sarah Josepha Hale Award. Some of Picoult’s most popular novels include *My Sister’s Keeper*, *Nineteen Minutes*, and *House Rules*.

259. Picoult is a member of The Authors Guild.

260. Picoult is the sole author of and owner or beneficial owner of the registered copyrights in twenty-seven (27) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Picoult Infringed Works”).

261. The registration information for the Picoult Infringed Works is contained in Exhibit A to this Complaint, at 7.

262. OpenAI unlawfully and willfully copied the Picoult Infringed Works and used them to “train” OpenAI’s LLMs without Picoult’s permission.

263. For example, when prompted, ChatGPT accurately generated summaries of several of the Picoult Infringed Works, including summaries for *Keeping Faith*, *Handle With Care*, and *Sing You Home*.

264. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Small Great Things*, one of the Picoult Infringed Works, and titled the infringing and unauthorized derivative “Small Great Things: Unfinished Business,” using the same characters from Picoult’s existing book.

265. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *My Sister's Keeper*, one of the Picoult Infringed Works, and titled the infringing and unauthorized derivative as “My Sister’s Legacy,” using the same characters from Picoult’s existing book.

266. When prompted, ChatGPT generated an accurate summary of the final chapter of *Change of Heart*, one of the Picoult Infringed Works.

267. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Picoult Infringed Works.

XIV. Plaintiff Preston

268. Preston is an author and journalist who has received awards for his writing, both in America and abroad, and previously taught writing at Princeton University. Some of Preston’s most popular novels include *Blasphemy*, *Impact*, and *The Codex*.

269. Preston is a member of The Authors Guild and past President of The Authors Guild Council.

270. Preston is the sole author of and owner or beneficial owner of the registered copyrights in six (6) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Preston Infringed Works”).

271. The registration information for the Preston Infringed Works is contained in Exhibit A to this Complaint, at 7.

272. OpenAI unlawfully and willfully copied the Preston Infringed Works and used them to “train” OpenAI’s LLMs without Preston’s permission.

273. For example, when prompted, ChatGPT accurately generated summaries of several of the Preston Infringed Works, including summaries for *Impact*, *Blasphemy*, and *The Codex*.

274. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Impact*, one of the Preston Infringed Works, and titled the infringing and unauthorized derivative “Unearthed Secrets,” using the same characters from Preston’s existing book.

275. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Codex*, one of the Preston Infringed Works, and titled the infringing and unauthorized derivative “The Codex: The Lost Dynasty,” using the same characters from Preston’s existing book.

276. When prompted, ChatGPT generated an accurate summary of the final chapter of *The Kraken Project*, one of the Preston Infringed Works.

277. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Preston Infringed Works.

XV. Plaintiff Robinson

278. Robinson is an award-winning author with a wide reach, having written six novels and three collections of short stories, whose fiction has appeared in internationally respected publications and whose books have been published internationally. Some of Robinson’s most popular novels include *Dawson’s Fall*, *Sparta*, and *Cost*.

279. Robinson is a member of The Authors Guild and a past President of The Authors Guild Council.

280. Robinson is the sole author of and owner or beneficial owner of the registered copyrights in eight (8) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Robinson Infringed Works”).

281. The registration information for the Robinson Infringed Works is contained in Exhibit A to this Complaint, at 7–8.

282. OpenAI unlawfully and willfully copied the Robinson Infringed Works and used them to “train” OpenAI’s LLMs without Robinson’s permission.

283. For example, when prompted, ChatGPT accurately generated summaries of several of the Robinson Infringed Works, including summaries of *Cost*, *Sparta* and *Dawson’s Fall*.

284. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Dawson’s Fall*, one of the Robinson Infringed Works, and titled the infringing and unauthorized derivative “Dawson’s Legacy,” using the same characters from Robinson’s existing book.

285. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Sparta*, one of the Robinson Infringed Works, and titled the infringing and unauthorized derivative “Homefront,” using the same characters from Robinson’s existing book.

286. When prompted, ChatGPT generated an accurate summary of the final chapter of *Sparta*, one of the Robinson Infringed Works.

287. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Robinson Infringed Works.

XVI. Plaintiff Saunders

288. Saunders is a professor in the English department at Syracuse University, who also writes best-selling books of fiction. Some of Saunders’ most popular works include the short story titled *Escape From Spiderhead*, a novel titled *Lincoln in the Bardo*, and a novella titled *The Brief and Frightening Reign of Phil*.

289. Saunders is a member of The Authors Guild.

290. Saunders is the sole author of and owner or beneficial owner of the registered copyrights in seven (7) written works of fiction, all or many of which OpenAI ingested and copied without permission (the “Saunders Infringed Works”).

291. The registration information for the Saunders Infringed Works is contained in Exhibit A to this Complaint, at 8.

292. OpenAI unlawfully and willfully copied the Saunders Infringed Works and used them to “train” OpenAI’s LLMs without Saunders’s permission.

293. For example, when prompted, ChatGPT accurately generated summaries of several of the Saunders Infringed Works, including summaries for *CivilWarLand in Bad Decline*, *Lincoln in the Bardo*, and *Tenth of December*.

294. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Fox 8*, one of the Saunders Infringed Works, and titled the infringing and unauthorized derivative “Fox 8 and the Hidden World,” using the same characters from Saunders’s existing book.

295. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Tenth of December*, one of the Saunders Infringed Works, and titled the infringing and unauthorized derivative “The Eleventh of December: A Continuation,” using the same characters from Saunders’s existing book.

296. When prompted, ChatGPT generated an accurate summary of the conclusion of *Escape From Spiderhead*, one of the Saunders Infringed Works.

297. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Saunders Infringed Works.

XVII. Plaintiff Turow

298. A best-selling author, Turow is a novelist and lawyer who is best known for setting his novels in fictional Kindle County's legal community. Some of Turow's most popular novels include *The Last Trial*, *Testimony*, and *Identical*.

299. Turow is a member of The Authors Guild and past President of The Authors Guild Council.

300. Turow is the sole author of and owner or beneficial owner of the registered copyrights in sixteen (16) written works of fiction, all or many of which OpenAI ingested and copied without permission (the "Turow Infringed Works").

301. The registration information for the Turow Infringed Works is contained in Exhibit A to this Complaint, at 8.

302. OpenAI unlawfully and willfully copied the Turow Infringed Works and used them to "train" OpenAI's LLMs without Turow's permission.

303. For example, when prompted, ChatGPT accurately generated summaries of several of the Turow Infringed Works, including summaries for *The Burden of Proof*, *Innocent*, and *Testimony*.

304. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *The Last Trial*, one of the Turow Infringed Works, and titled the infringing and unauthorized derivative "Echoes of Judgment," using the same characters from Turow's existing book.

305. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Pleading Guilty*, one of the Turow Infringed Works, and titled the infringing and unauthorized sequel "Redemption's Price," using the same characters from Turow's existing book.

306. When prompted, ChatGPT generated an accurate summary of the final chapter of *Ordinary Heroes*, one of the Turow Infringed Works.

307. ChatGPT could not have generated the material described above if OpenAI's LLMs had not ingested and been "trained" on the Turow Infringed Works.

XVIII. Plaintiff Vail

308. Rachel Vail is an award-winning American author who primarily authors children's and young adult books. Some of Vail's most popular novels include *Ever After*, *Unfriended*, and *Justin Case: School, Drool, and Other Daily Disasters*.

309. Vail is a member of The Authors Guild and a member of The Authors Guild Council.

310. Vail is the sole author of and owner or beneficial owner of the registered copyrights in twenty-four (24) written works of fiction, all or many of which OpenAI ingested and copied without permission (the "Vail Infringed Works").

311. The registration information for the Vail Infringed Works is contained in Exhibit A to this Complaint, at 8–9.

312. OpenAI unlawfully and willfully copied the Vail Infringed Works and used them to "train" its LLMs without Vail's permission.

313. For example, when prompted, ChatGPT accurately generated summaries of several of the Vail Infringed Works, including summaries for *If We Kiss*, *A Is For Elizabeth*, and *Not That I Care*.

314. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Bad Best Friend*, one of the Vail Infringed Works, and titled the infringing and unauthorized derivative "Redeeming Friendship," using the same characters from Vail's existing book.

315. When prompted, ChatGPT generated an infringing, unauthorized, and detailed outline for the next purported installment of *Do-Over*, one of the Vail Infringed Works, and titled the infringing and unauthorized derivative “Do-Over: Second Chances,” using the same characters from Vail’s existing book.

316. When prompted, ChatGPT generated an accurate summary of the final chapter of *Daring to be Abigail*, one of the Vail Infringed Works.

317. ChatGPT could not have generated the material described above if OpenAI’s LLMs had not ingested and been “trained” on the Vail Infringed Works.

XIX. Others Similarly Situated

318. The above allegations illustrate certain specific ways in which OpenAI’s LLM “training” has infringed Plaintiffs’ copyrights and has injured or may injure the value of their works. But OpenAI has engaged in a systematic course of mass-scale copyright infringement that violates the rights of all working fiction writers and their copyright holders equally, and threatens them with similar, if not identical, harm.

CLASS ALLEGATIONS

I. Class Definitions

319. Plaintiffs bring this action as Proposed Class Representatives for and on behalf of the Proposed Class and Proposed Class Members, as defined below, under Federal Rules of Civil Procedure 23(b)(2), 23(b)(3), and 23(c)(4).

320. The Proposed Class is defined as follows:

All natural persons in the United States who are the sole authors of, and sole legal or beneficial owners of Eligible Copyrights in, one or more Class Works; and all persons in the United States who are the sole legal or beneficial owners of Eligible Copyrights in one or more Class Works held by literary estates.

321. Class Works are defined as follows:

Any work of fiction that has sold at least 5,000 copies and the text of which has been, or is being, used by Defendants to “train” one or more of Defendants’ large language models.

322. Eligible Copyrights are defined as follows:

Any copyright that was registered with the United States Copyright Office before or within five years after first publication of the work, and whose effective date of registration is either within three months after first publication of the work or before Defendants began using the work to “train” one or more of Defendants’ large language models.

323. Excluded from the class definitions above are Defendants; Defendants’ co-conspirators, aiders and abettors, and members of their immediate families; Defendants’ corporate parents, subsidiaries, and affiliates; Defendants’ directors, officers, employees, and other agents, as well as members of their immediate families; and any judge who may preside over this action, the judge’s staff, and members of their immediate families.

II. Rules 23(a) and 23(g)

324. The Proposed Class is sufficiently numerous because it is estimated to have tens of thousands of members.

325. The identities of the Proposed Class Members are objectively ascertainable because Defendants know, and can produce in discovery, which texts they used to “train” their large language models; and because information regarding copyright ownership, copyright registration, and book sales is determinable from public or other objective sources and measures.

326. The Proposed Class Representatives’ claims are typical of the claims of the Proposed Class because their copyrights were infringed in materially the same way and their interests in preventing future infringement and redressing past infringement are materially the same.

327. The Proposed Class Representatives will adequately represent the Proposed Class, and Plaintiffs' counsel are experienced, knowledgeable, well resourced, and will zealously and faithfully represent Plaintiffs and the Proposed Class.

328. There are questions of law or fact common to the Proposed Class, including

- a. whether Defendants copied Plaintiffs' and Proposed Class Members' copyrighted works in "training" their LLMs;
- b. whether Defendants' copying of Plaintiffs' and Proposed Class Members' copyrighted works constitutes direct, vicarious, or contributory infringement under the Copyright Act; and
- c. whether Defendants' copying of Plaintiffs' and Proposed Class Members' copyrighted works was willful.

III. Rule 23(b)

329. Defendants have acted on grounds common to Plaintiffs and the Proposed Class by treating all Plaintiffs' and Proposed Class Members' works equally, in all material respects, in their LLM "training."

330. Common questions of liability for infringement predominate over any individualized damages determinations as may be necessary. To decide liability, the Court will necessarily apply the same law to the same conduct, which Defendants engaged in indiscriminately with respect to all Plaintiffs and all Proposed Class Members.

331. Further, to the extent Plaintiffs elect to pursue statutory rather than actual damages before final judgment, the damages inquiry will likewise be common, if not identical, across Plaintiffs and Proposed Class Members.

332. A class action is superior to any individual litigation of Plaintiffs' and Proposed Class Members' claims. Proposed Class Members have little interest, distinct from Plaintiffs'

and other Proposed Class Members’, in prosecuting individual actions. It would waste judicial resources to decide the same legal questions repeatedly, thousands of times over, on materially indistinguishable facts. The Proposed Class presents no special manageability problems.

IV. Rule 23(c)(4)

333. In the alternative to certification under Rule 23(b)(3), common questions predominate within the determination of liability for infringement, and therefore the issue of liability may be separately certified for class treatment even if the entire action is not.

CLAIMS TO RELIEF

COUNT I: DIRECT COPYRIGHT INFRINGEMENT (17 U.S.C. § 501)

**On Behalf of Plaintiffs and the Proposed Class
Against Defendant OpenAI OpCo LLC**

334. Plaintiffs incorporate and reallege paragraphs 1 through 318 above.

335. Plaintiffs and Proposed Class Members are the rightful and lawful legal or beneficial owners of the copyrights in and to their literary works.

336. Plaintiffs’ and Proposed Class Members’ works are original to their authors and are fixed in tangible mediums of expression as literary works under 17 U.S.C. § 102(a)(1).

337. Plaintiffs and Proposed Class Members have duly and timely registered their copyrights in their works with the U.S. Copyright Office.

338. Plaintiffs and Proposed Class Members are legal or beneficial owners of the exclusive right to reproduce their copyrighted works in copies under 17 U.S.C. § 106(1), as well as the right to refrain from such reproduction.

339. Defendant OpenAI OpCo LLC had access to Plaintiffs’ and Proposed Class Members’ copyrighted works, including by way of the various unauthorized datasets discussed above.

340. Defendant OpenAI OpCo LLC violated Plaintiffs’ and Proposed Class Members’ exclusive rights by reproducing their copyrighted works in copies for the purpose of “training” their LLMs and ChatGPT.

341. Defendant OpenAI OpCo LLC’s violation of Plaintiffs’ and Proposed Class Members’ exclusive right was willful because Defendant OpenAI OpCo LLC knew the datasets on which it “trained” its large language models contained copyrighted works.

COUNT II: VICARIOUS COPYRIGHT INFRINGEMENT

On Behalf of Plaintiffs and the Proposed Class Against Defendants OpenAI Inc. and OpenAI GP LLC

342. Plaintiffs incorporate and reallege paragraphs 1 through 318 above.

343. Defendants OpenAI Inc. and OpenAI GP LLC had the right and ability to control the direct infringement alleged in Count I because Defendant OpenAI Inc. fully controls Defendant OpenAI GP LLC, and Defendant OpenAI GP LLC fully controls Defendant OpenAI OpCo LLC, according to the corporate structure outlined above.

344. Defendants OpenAI Inc. and OpenAI GP LLC have a direct financial interest in the direct infringement alleged in Count I because they benefit from the profits and investments generated by Defendant OpenAI OpCo LLC’s infringing activities.

345. Defendants OpenAI Inc. and OpenAI GP LLC are vicariously liable for the direct infringement alleged in Count I.

COUNT III: CONTRIBUTORY COPYRIGHT INFRINGEMENT

On Behalf of Plaintiffs and the Proposed Class Against Defendants OpenAI LLC, OpenAI Global LLC, OAI Corporation LLC, OpenAI Holdings LLC, OpenAI Startup Fund I LP, OpenAI Startup Fund GP I LLC, and OpenAI Startup Fund Management LLC (Collectively, “Other OpenAI Defendants”), and Defendant Microsoft

346. Plaintiffs incorporate and reallege paragraphs 1 through 318 above.

347. Microsoft and the Other OpenAI Defendants materially contributed to and directly assisted in the direct infringement alleged in Count I by funding the direct infringement by way of capital, technology, personnel, and other resources; controlling or managing the property or other assets with which the direct infringement was accomplished; or providing business, legal, strategic, or operational guidance to accomplish the direct infringement.

348. Microsoft and the Other OpenAI Defendants knew or had reason to know of the direct infringement alleged in Count I, because Microsoft and the Other OpenAI Defendants share management personnel and operational plans with Defendant OpenAI OpCo LLC and are fully aware of the capabilities of their own product and the materials upon which it was “trained,” including known caches of pilfered copyrighted works.

349. Defendants are contributorily liable for the direct infringement alleged in Count I.

PRAYER FOR RELIEF

350. Plaintiffs, on behalf of themselves and all others similarly situated, pray for the following relief:

- a. Certification of this action as a class action under Federal Rule of Civil Procedure 23;
- b. Designation of Plaintiffs as class representatives;
- c. Designation of Plaintiffs’ counsel as class counsel;
- d. An injunction prohibiting Defendants from infringing Plaintiffs’ and class members’ copyrights, including without limitation enjoining Defendants from using Plaintiffs’ and class members’ copyrighted works in “training” Defendants’ large language models without express authorization;
- e. An award of actual damages to Plaintiffs and class members;

f. An award of Defendants' additional profits attributable to infringement to Plaintiffs and class members;

g. An award of statutory damages up to \$150,000 per infringed work to Plaintiffs and class members, in the alternative to actual damages and profits, at Plaintiffs' election before final judgment;

h. Reasonable attorneys' fees and costs, as allowed by law;

i. Pre-judgment and post-judgment interest, as allowed by law; and

j. Such further relief as the Court may deem just and proper.

JURY DEMAND

351. Plaintiffs demand a trial by jury as to all issues so triable.

Dated: Dec. 4, 2023

Respectfully submitted,

/s/ Rachel Geman

Rachel Geman
LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
250 Hudson Street, 8th Floor
New York, NY 10013-1413
Telephone: 212.355.9500
rgeman@lchb.com

Ian R. Bensberg (*pro hac vice* submitted)
LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
275 Battery Street, 29th Floor
San Francisco, CA 94111-3339
Telephone: 415.956.1000
rstoler@lchb.com
ibensberg@lchb.com

Wesley Dozier (*pro hac vice* submitted)
LIEFF CABRASER HEIMANN & BERNSTEIN, LLP
222 2nd Avenue, Suite 1640
Nashville, TN 37201
Telephone: 615.313.9000
wdozier@lchb.com

/s/ Scott Sholder

Scott J. Sholder
CeCe M. Cole
COWAN DEBAETS ABRAHAMS & SHEPPARD LLP
41 Madison Avenue, 38th Floor
New York, New York 10010
Telephone: 212.974.7474
nwolff@cdas.com
kswezey@cdas.com
ssholder@cdas.com
ccole@cdas.com

Attorneys for Plaintiffs and the Proposed Class